DOCUMENT RESUME

ED 290 801 TM 870 663

AUTHOR Halpin, Gerald; Simpson, Robert
TITLE Psychometric Characteristics Associated with Changes on the Passage Comprehension Test of the Woodcock Reading Mastery Tests.
PUB DATE Nov 86
NOTE 16p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Memphis, TN, November 19-21, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adolescents; Analysis of Variance; *Behavior Problems; Raw Scores; *Reading Tests; Regression (Statistics); Scoring Formulas; Student Motivation; *Test Reliability
IDENTIFIERS *Passage Comprehension Test; Woodcock Reading Mastery Test

ABSTRACT

Forty adolescent subjects with behavior problems at home or at school were administered the Woodcock Reading Mastery Tests. Subjects ranged in age from 12 to 18, and included 25 males and 15 females. The Passage Comprehension Test for each subject was rescored using three different ceiling criteria: (1) five errors in six consecutive responses, (2) five errors in seven consecutive responses, and (3) three consecutive errors. The three scoring methods were found to be equivalent to the scoring method recommended in the Woodcock Manual in terms of internal consistency reliability. Problems were observed with student frustration resulting in discouragement and diminished motivation and excessive time spent in test administration. Strong positive correlations were found between raw scores obtained by the four scoring methods and between each group of raw scores and an external measure. Analysis of variance revealed statistically significant differences between the raw score means. Bivariate regression equations were provided to allow raw scores obtained with each of the three scoring methods to be adjusted so they can be used to obtain percentile ranks and equivalent scores. Two alternative scoring methods reduced to a statistically significant degree the mean number of items administered with no loss in reliability or validity. Practical recommendations for diagnosticians are provided. (Author/RR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED290801

PSYCHOMETRIC CHARACTERISTICS ASSOCIATED WITH CHANGES ON THE PASSAGE COMPREHENSION TEST OF THE WOODCOCK READING MASTERY TESTS

Gerald Halpin and Robert Simpson

Auburn University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Gerald Halpin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running Head: ALTERING CEILING CRITERION

Paper presented at the meeting of the Mid-South Educational Research Association, Memphis, Tennessee, November 1986.

TM 870 663

BEST COPY AVAILABLE

Abstract

Forty adolescent subjects were administered the Woodcock Reading Mastery Tests. The Passage Comprehension Test for each subject was rescored using three different ceiling criteria: (a) five errors in six consecutive responses, (b) five errors in seven consecutive responses and (c) three consecutive errors. The three scoring methods were found to be equivalent to the scoring method recommended in the Woodcock Manual in terms of internal consistency reliability. Strong positive correlations were found between raw scores obtained by the four scoring methods and between each group of raw scores and an external measure. Analysis of variance revealed statistically significant differences between the raw score means. Bivariate regression equations were provided to allow raw scores obtained with each of the three scoring methods to be adjusted so they can be used to obtain percentile ranks and equivalent scores. Two alternative scoring methods reduced to a statistically significant degree the mean number of items administered with no loss in reliability or validity. Practical recommendations for diagnosticians are provided.

The Effects of Altering the Ceiling Criterion on the Passage Comprehension Test of the Woodcock Reading Mastery Tests

The Woodcock Reading Mastery Tests (Woodcock, 1973) (WRMT) are a battery of five individually administered reading tests designed to measure achievement in letter identification, word identification, word attack, word comprehension and passage comprehension. According to the test manual, all five tests can be administered in 20 to 30 minutes. Given the requirements of each test, administration of the Passage Comprehension Test often requires the greatest amount of time because, for each item, the subject must read a given passage and then supply an appropriate word that has been omitted.

The subject must correctly respond to five consecutive items to establish a basal level and must incorrectly respond to five consecutive items to establish a ceiling level. All items between the basal and ceiling must be administered. Two potential problems may occur when such a strict ceiling criterion is employed. The first potential problem is student frustration, resulting in discouragement and diminished motivation. The second is an undue amount of time expended in administering the test.

Consider the subject who misses four items, answers the next correctly, misses three items, answers the next correctly, and then misses five consecutive items. In responding to the test, the subject has answered incorrectly 12 of the final 14 items administered. Sometimes the subject appears to be aware that the items are difficult, or that he/she is answering many items incorrectly. The subject may even admit to guessing. Repeatedly missing items could lead to frustration, discouragement and diminished motivation to do well. The result could be deflated scores on tests subsequently administered in the same battery.

The second potential problem concerns time expended in administering the Passage Comprehension Test. Instructions in the WRMT Manual indicate that the subject should respond "fairly promptly" and that the examiner should not wait "an undue length of time" for the subject to respond to each test item. No specific response time is suggested, and deciding what is an appropriate response time is apparently left to the discretion of the examiner. Nevertheless, when the ceiling level is being approached, the subject may tend to require more response time per item because of increasing item difficulty and still miss many items before a ceiling is reached. The result can be an inordinate amount of time required to administer the Passage Comprehension Test.

Given the potential problems of student frustration, diminished motivation and undue expenditure of time, the present study was designed to determine the effects of employing less stringent ceiling criteria when administering the Passage Comprehension Test of the WRMT. If less stringent criteria have a negligible effect on psychometric characteristics while, at the same time, result in fewer items administered to establish a ceiling level, then a logical case can be made for employing less stringent criteria. A literature review yielded no reported studies in which researchers investigated the feasibility of altering the scoring criterion for any of the Woodcock Reading Mastery Tests.

In the WRMT Manual, a ceiling criterion of three consecutive errors is recommended as one way to conserve testing time. Other tests, such as the Peabody Individual Achievement Test (Dunn & Mardwardt, 1970), employ a ceiling criterion of five errors in seven consecutive responses. In the present study, the effects of three alternate ceiling criteria were investigated. The three criteria were (a) five errors in six consecutive responses, (b) five errros in seven consecutive responses and (c) three consecutive errors.

Specific research questions for this investigation were as follows:

1. What are the effects on internal consistency of the Passage Comprehension Test when the three alternate scoring methods are employed?
2. Does the rank order of the raw scores change as a function of scoring method?
3. What effect does altering the scoring method have on the external validity of the Passage Comprehension Test?
4. Does varying the scoring method result in raw scores that systematically differ from raw scores obtained using the scoring method recommended in the WRMT Manual?
5. Is there a statistically significant difference between the mean number of test items administered using each scoring criterion?
6. What corrections are necessary to make scores obtained using the alternate scoring methods comparable to raw scores obtained using the method recommended in the WRMT Manual?

Method

Subjects

Subjects of this study were 40 adolescents who were referred to a diagnostic and evaluation center because of behavior problems either at home or at school. The center accepted

referrals only for children who were not considered to have serious emotional difficulties. The subjects ranged in age from 12 years 2 months to 18 years 0 months. The sample included 25 males and 15 females. There were 45 whites and 5 blacks. The measured intelligence quotients ranged from 73 to 115 with a mean of 92 and standard deviation of 9.76.

Procedure

As part of a standard educational and intellectual assessment battery, all children in the diagnostic center were administered the Woodcock Reading Mastery Tests, Form A. Forty children who had already been tested were randomly selected to serve as subjects, and their corresponding WRMT protocols were retrieved from case files. On each protocol, the Passage Comprehension Test was rescored, employing three different ceiling criteria. The three criteria were: (a) five errors in six consecutive responses (Scoring Method 2), (b) five errors in seven consecutive responses (Scoring Method 3), and (c) three consecutive errors (Scoring Method 4). Use of the ceiling criterion described in the WRMT Manual was Scoring Method 1. Since establishment of a basal usually involves administering relatively easy items, thereby requiring less time, the basal criterion of five consecutive responses was employed across all scoring methods.

The corrected Kuder-Richardson 21 formula (Wilson, Downing & Ebel, 1977) was used to calculate the internal consistency reliability of the Passage Comprehension Test utilizing each of the four scoring methods. Pearson product moment correlation coefficients were computed between raw scores obtained using the three alternate scoring methods and raw scores obtained using the scoring method described in the WRMT Manual. To investigate the effect of altering the ceiling criterion on the validity of the Passage Comprehension Test, Pearson product moment correlation coefficients were calculated to determine the degree of relationship between raw scores obtained by employing the four ceiling criteria and an external criterion measure. The external criterion measure was the full scale intelligence quotient obtained by administering the Wechsler Intelligence Scale for Children-Revised (Wechsler, 1974) or the Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981). The appropriate intelligence test was determined by the age of the subject.

One factor repeated measures analysis of variance procedures were used to compare the mean raw scores obtained and mean number of test items administered, utilizing each of the four scoring methods. Tukey's a posteriori multiple comparison procedure was used to further identify mean differences. Simple bivariate regression equations were generated to predict raw scores

obtained using procedures contained in the WRMT Manual from raw scores obtained using the three alternate scoring methods.

Results

In Table 1 are presented three sets of correlation coefficients. The first four correlation coefficients are the corrected Kuder-Richardson 21 internal consistency estimates computed using each of the four scoring methods. Perusal of Table 1 reveals that, within rounding error, the four estimates are the same. The second set of correlation coefficients

---

Insert Table 1 about here

---

indicates the degree of relationship between raw scores obtained from Scoring Methods 2, 3 and 4 respectively with Scoring Method 1 (the method recommended in the WRMT Manual). The correlations ranged from .98 to .99. The final set of correlation coefficients indicates the degree of relationship between full scale intelligence quotients and raw scores obtained when using each of the four scoring criteria.

The mean raw scores and mean number of items administered for each of the four ceiling criteria, along with the corresponding standard deviations, are presented in Table 2.

Insert Table 2 about here

Results of the one factor repeated measures analysis of variance revealed a statistically significant difference between the four raw score means obtained when using the various scoring methods, F (3, 117) = 12.11, $p < .001$. Utilizing Tukey's test for comparing means obtained under the various scoring methods, a significant difference was obtained between the mean from Method 1 (M = 50.33) and the means from Scoring Methods 3 (M = 49.08) and 4 (M = 48.23) respectively. There was not a significant difference obtained between the mean from Method 1 and the mean from Method 2 (M = 49.83). Analysis of variance also revealed a statistically significant difference between the means of the number of items administered using each scoring criterion, F (3, 117) = 30.06, $p < .001$.

A bivariate regression equation, $Y' = AX + B$, was derived for each of the three alternate scoring methods so that raw scores obtained using Method 1 could be predicted from raw scores obtained using Methods 2, 3 and 4 respectively. The three regression equations are presented below:

for Method 2: $Y' = .99\ X + .80$

for Method 3: $Y' = .99\ X + 1.80$

for Method 4: $Y' = 1.01\ X + 1.56$

Discussion

Data presented in Table 1 indicate that the internal consistency of the Passage Comprehension Test remains high across all four scoring methods. Also, the correlations between raw scores for each method are so strongly positive that they indicate almost no change in the rank order of scores from one method to another. The correlation coefficients indicating the degree of relationship between raw scores and an external measure, full scale intelligence quotient, vary from each other by no more than .04 correlational points.

Since the raw scores obtained under each scoring method were found to be virtually uninfluenced in terms of reliability and validity, they can be made usable by applying the simple bivariate regression equations presented in the Results Section. The regression equations provide an adjustment factor so that raw scores obtained with each of the three alternate scoring methods can be used to obtain equivalent scores and percentile ranks from the WRMT norm tables.

In practical application, if Scoring Method 2 (five errors in six responses) is employed, one point should be added to the

raw score before consulting the norm tables. If Methods 3 (five errors in seven responses) or 4 (three consecutive errors) are used, 2 points should be added to the raw score.

Examination of data in Table 2 reveals that the mean number of items administered decreases by 10%, 18% and 26% for Scoring Methods 2, 3 and 4 respectively. Since the four scoring methods are equivalent in terms of internal consistency, external validity, and rank ordering of raw scores and, given the substantial reduction in the percentage of mean number of items administered, Scoring Method 3 or 4 is recommended when the examiner wants to reduce the time required for test administration. This more efficient use of testing time is likely to result in a substantially less frustrating testing situation since the student must miss fewer items before the testing ceiling is established. Reduced student frustration could lead to better performance during the remainder of the testing session and, therefore, a more accurate assessment of the student's capabilities.

Although results of the present study should be verified with a larger number of students, representing various age levels, test users should feel relatively confident that using the shortened versions of the Passage Comprehension Test will result in little loss in psychometric quality.

References

Dunn, L.M. & Markwardt, F. C. (1970). Peabody Individual Achievement Test. Circle Pines, MN: American Guidance Service.

Wechsler, D. (1981). Wechsler Adult Intelligence Scale-Revised: Manual. New York: Psychological Corporation.

Wechsler, D. (1974). Wechsler Intelligence Scale for Children-Revised: Manual. Cleveland: Psychological Corporation.

Wilson, P. A., Downing, S. M., & Ebel, R. L. (1977). An empirical adjustment of the Kuder-Richardson 21 Reliability Coefficient to better estimate the Kuder-Richardson 20 Coefficient. Unpublished manuscript.

Woodcock, R. (1973). Woodcock Reading Mastery Tests: Manual. Circle Pines, MN: American Guidance Service.

Table 1

Correlations: Internal Consistency, Intramethod, and External Criterion

Scoring Method	Ceiling Criterion	Internal Consistency Reliability	Intramethod Correlations	Correlations With IQ
1	5/5	.93	----	.42*
2	5/6	.93	.99*	.44*
3	5/7	.93	.99*	.43*
4	3/3	.93	.98*	.40*

* $p < .01$

Table 2

Raw Score Means, Mean Number of Items Administered and Standard Deviations for each Ceiling Criterion

Ceiling Criterion	Mean Raw Score	Standard Deviation	Mean # Items Administered	Standard Deviation
5/5	50.33	15.61	28.58	11.62
5/6	49.83	15.65	25.70	10.17
5/7	49.08	15.58	23.55	9.90
3/3	48.23	15.15	21.25	10.31